Intersection of Transportation Infrastructure and Displacement

December 12, 2023

Alexander Heger, Raviv Zait, Kwan Wing Tuet, Siddharth Bookinkere

Abstract

The iconic 1989 sport-fantasy film, *Field of Dreams*, is perhaps most known for an adapted quote from its script–"If you build it, [they] will come." But, some 1,200 miles east of rural Iowa, what happens when infrastructure intended to assuage transportation challenges is installed in some of Boston's poorest neighborhoods? Could building bike lanes actually force vulnerable populations out of their homes? This report seeks to answer this question through causal analysis of bike infrastructure development and displacement in Boston's Dorchester and Roxbury neighborhoods. Additionally, we look at areas surrounding the Forest Hills Massachusetts Bay Transportation Authority (MBTA) Station in Jamaica Plain to consider the same effects of new greenway development. We use data from the United States Census' American Community Survey's Five-Year Estimates and the City of Boston's Boston Bikes Data webpage to find out if there are (1) correlations between bike lane development and the factors leading to displacement, (2) different effects as a result of different types of bike infrastructure, (3) common demographic and socioeconomic characteristics of neighborhoods where bike lanes are built, and (4) if these characteristics are shared between bike lanes and greenways?

Table of Contents

1. Introduction

The intersection of transportation infrastructure and displacement is a complex and nuanced topic that involves examining causation and correlation between various factors. Currently, research often looks at indicators such as the introduction of bike lanes or new businesses like Whole Foods as last indicators of displacement. However, there is a significant gap in our understanding, and the complexity of this issue presents challenges in data analysis.

LivableStreets Alliance, the client for this project, is a 501(c)(3) nonprofit located in Cambridge, Massachusetts. Their mission is to advocate for transportation solutions that are safe, affordable, and enjoyable by dismantling geographical and social barriers that divide communities. Two forms of infrastructure for which LivableStreets advocates are bike lanes and greenways. These projects are met with resistance in some communities, where residents fear that development of bike lanes and greenways may displace them through increased costs of living. Therefore, displacement is a critical concern for LivableStreets, who seeks to understand the factors leading up to the introduction of amenities like bike lanes, bus lanes, and train access.

National data indicates a correlation between displacement and the presence of train stops, but painted bus, bike lanes, parks and greenways are not analyzed as triggers. Delving into the specifics of displacement triggered by factors like bus lanes and bike infrastructure in the Boston Metro area could yield valuable insights to LivableStreets. Understanding whether causation exists between these factors and displacement could inform the implementation of policies to prevent or mitigate it. Conversely, if no causation is found, this information can be used to educate the public about the actual dynamics at play.

2. Defining Displacement

Before displacement can be measured with regard to transportation infrastructure, it must be defined. This section outlines our conversations with LivableStreets to come to a mutual understanding of what displacement looks like in these Boston neighborhoods.

2.1. Fears of Displacement in Suffolk County

Our project partners at LivableStreets Alliance noted a concern among residents of neighborhoods where transportation infrastructure was being developed. They often heard in community forums that new infrastructure would increase their costs of living such that they would be forced out of their homes, unable to pay to stay. One instance of this was documented by GBH News in 2020, where residents confided that the changes in their South Mattapan community caused by a new MBTA stop were causing residents to be displaced (McKim). These changes included land being purchased by real estate developers who designed expansive apartment complexes whose rents required smaller-scale landlords to increase their own prices. This makes elderly residents and long-term renters particularly vulnerable as a result of their potentially fixed incomes, and the lack of ability to negotiate lease terms. These fears are not unique to the Boston area, and can be found in many disadvantaged neighborhoods and communities around the world (see for example The Guardian, 2016).

Beyond economic changes, other residents feared that the new station would spur gentrification and the racial whitening of their neighborhood at the expense of its predominantly Black community. A resident is quoted as saying, “This train is not for us. It's for the white people who want to move back from the suburbs into the city who left in the first place when we moved here” (McKim). Gentrification, the sociological phenomenon whereby wealthy newcomers move into, and displace the inhabitants of, a poorer urban neighborhood, is often characterized in the United States by white communities replacing Black. This fact is tied to the United States’ history of systematic oppression of Black and African American populations such that they are forced into lower wealth communities.

2.2. Observable Effects of Displacement

The experience of LivableStreets combined with testimonies similar to the above revealed to us four key factors that are observed when displacement occurs. These are changes in (1) income, (2) property value, (3) racial composition of communities, and (4) rates of home ownership. As documented in the GBH article, the fear of impending gentrification in South Mattapan was observed through real estate developers’ building apartments for white collar commuters (factor 1), who could afford higher rents (factor 2), who were likely white (factor 3), and whose presence would mark an increase in rented housing units (factor 4). Rates of homeownership can also indicate how dependent a community is on temporary housing, which can be linked to lower income populations. This is evident in parts of Dorchester and Roxbury, our two primary communities of interest.

2.3. Looking for Displacement in Census Data

With the relevant factors of displacement identified for our project scope, we turned to the United States Census in search for data. We started our search by considering different geographies used to organize Census data. These range from the whole of the United States at the most broad, to Census Block Groups at the most narrow, as shown in Figure 2.1. We chose Census Tracts, two categorizations more broad than Block Groups, due to their likeness in size to Boston neighborhoods. Census Tracts are meant to capture information on up to 4,000 residents on average (U.S. Census). One complicating factor in our choice of organizing our data around Census Tracts is that Census Tracts, while relatively static, changed in 2020. This meant that our data from 2020 and 2021 had to be attributed to 2010 Census Tracts in our data cleaning so information from these more recent years was not lost.

Table 2.1 - Selected Census Data Tables

Displacement Factor	Selected Census Table
Income	S1901 - Income in the Past 12 Months (in 2021 Inflation-Adjusted Dollars)
Race	DP05 - ACS Demographic and Housing Estimates
Home Ownership	B25026 - Total Population in Occupied Housing Units by Tenure by Year Householder Moved into Unit
Property Value	DP04 - Selected Housing Characteristics

We then turned our attention to the specific data tables within the Census that would provide enough information to answer our analytical questions without being too lengthy or complex. Each data table is a subcomponent of a survey, and we selected the American Community Survey for the versatility of its pentannual estimates of many social factors. As is a theme with the Census, there is a broad range of selections to make when choosing data tables. We landed on four which each highlighted one of the aforementioned factors of displacement, shown in Table 2.1. From there, we then had to clean and prepare the data for analysis in Python.

3. Data Collection and Processing

While the United States Census Bureau has an application programming interface (API) for developers and researchers, we opted to download comma separated value (CSV) files of the tables we selected using Census Bureau Tables tool due to its ease of use and the short duration of this project. Once each of the tables was downloaded, it was a relatively simple process to clean and merge them in Python. The same could have been done in Excel, but this technique would have been far more time consuming and prone to error.

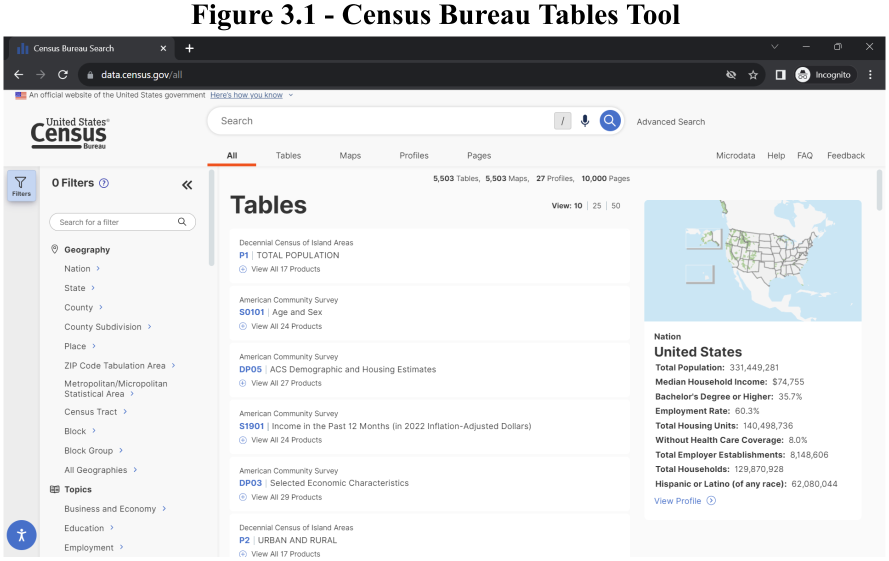
Since social factors of displacement are the independent variables of this study, we also needed our independent variable–bike infrastructure development. For this data, we turned to a repository from the Boston Transportation Department.

3.1. American Community Survey Five-Year Estimates

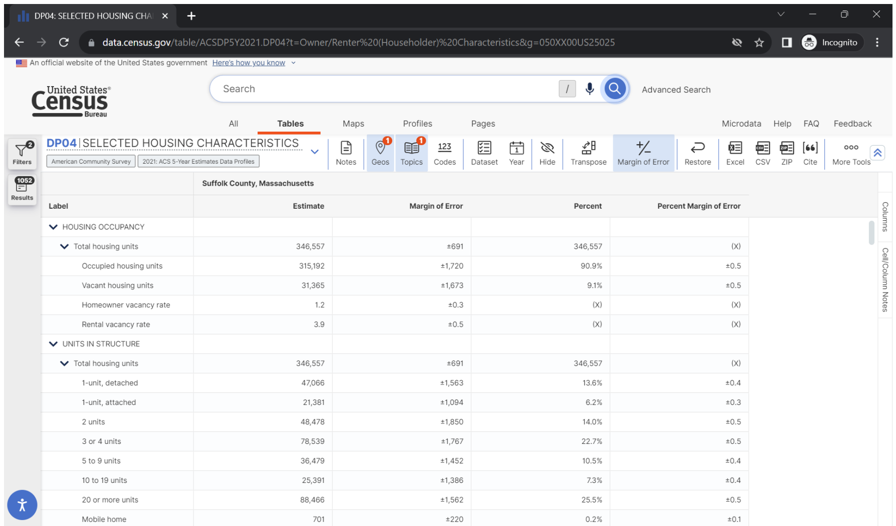
The American Community Survey (ACS) is a detailed housing and demographic information survey that is released each year by the United States Census. The data are primarily conveyed as estimates based on surveys and public records collected within the span of 1 or 5 years. Given that we would only be looking at bike infrastructure data from 2008 to 2023, we selected the ACS 5-year estimates to maximize the demographic and housing information taken into consideration to form each estimate.

Figure 3.1 - Census Bureau Tables Tool

(1)



(2)



Label	Estimate	Margin of Error	Percent	Percent Margin of Error
HOUSING OCCUPANCY				
Total housing units	346,557	±691	346,557	(X)
Occupied housing units	315,192	±1,720	90.9%	±0.5
Vacant housing units	31,365	±1,673	9.1%	±0.5
Homeowner vacancy rate	1.2	±0.3	(X)	(X)
Rental vacancy rate	3.9	±0.5	(X)	(X)
UNITS IN STRUCTURE				
Total housing units	346,557	±691	346,557	(X)
1-unit, detached	47,066	±1,563	13.6%	±0.4
1-unit, attached	21,381	±1,094	6.2%	±0.3
2 units	48,478	±1,850	14.0%	±0.5
3 or 4 units	78,539	±1,767	22.7%	±0.5
5 to 9 units	36,479	±1,452	10.5%	±0.4
10 to 19 units	25,391	±1,386	7.3%	±0.4
20 or more units	88,466	±1,562	25.5%	±0.5
Mobile home	701	±220	0.2%	±0.1

At top, the Census Bureau Tables homepage; at bottom, a sample table for Suffolk County

Another advantage of the ACS 5-year estimates is that annual data is available at the Census Tract level and below, which was crucial for our year-over-year comparison of Boston neighborhoods. Our analysis comprised 47 distinct Census Tracts across Dorchester, Roxbury, and the Forest Hills area of Jamaica Plain.

In order to query the data tables we selected for the ACS 5-year estimates (Table 2.1) we used the Census Bureau Tables tool on the data.census.gov webpage, as shown in Figure 3.1. Within this tool, users can filter by geography, topic, survey, year, or code on the left panel of the webpage and results will populate to the right. Users have some 5,503 tables and maps from which to choose. When viewing each table online, users are limited in their options to interact with the data, but tables are easy to download to CSV, Excel, or a compressed folder.

Our data cleaning process involved transposing each data table so the rows represented Census Tracts and the columns were demographic information. Due to the tables' organization by year on the Census website, we had to download a different version of the table for each year, then merge them into one combined table in Python. From here, we cleaned string values to remove unnecessary characters, isolated only the estimates for each column, and renamed the columns to shorthand forms that were easier to include in our analysis. The complete code cell that we used for cleaning ACS Demographic and Housing Estimates from 2010 to 2021 is included in Figure 3.2.

Figure 3.2 - Cleaning Process for Race Data Table

```python
# Define a list of DataFrame names from d10 to d21
data_frame_names = ['d10', 'd11', 'd12', 'd13', 'd14', 'd15', 'd16', 'd17', 'd18', 'd19', 'd20', 'd21']

for name in data_frame_names:
    # Load the DataFrame from a CSV file
    df = pd.read_csv(data_folder + '20{}_demo.csv'.format(name[1:]), index_col=False)

    # Transpose the DataFrame
    df = df.transpose()

    # Extract the first row as column labels
    new_columns = df.iloc[0]

    # Set the first row as the column labels
    df = df.iloc[1:]
    df = df.set_axis(new_columns, axis=1)

    # Reset the index
    df = df.reset_index()
    df.columns.names = ['']

    # Rename the 'index' column to 'tract'
    df = df.rename(columns={"index": "tract"})

    # Split 'tract' column into 'tract' and 'info'
    df[['tract', 'info']] = df['tract'].str.split(',', 1, expand=True)

    # Clean the 'info' and 'tract' columns
    df['info'] = df['info'].str.replace('Suffolk County, Massachusetts!!', '', regex=True)
    df['tract'] = df['tract'].str.replace('Census Tract ', '', regex=True)

    # Remove commas
    df = df.replace(',', '', regex=True)

    # Remove spaces from column names
    df.columns = df.columns.str.strip()

    # Define columns to keep
    cols_to_keep = ['tract', 'Total population', 'Male', 'Female', 'White',
                    'Black or African American', 'American Indian and Alaska Native',
                    'Asian', 'Native Hawaiian and Other Pacific Islander', 'Some other race',
                    'Hispanic or Latino (of any race)', 'info', 'Median age (years)']

    # Drop columns not in the 'cols_to_keep' list
    df = df.drop([col for col in df.columns if col not in cols_to_keep], axis=1)

    # Remove duplicated columns
    df = df.loc[:, ~df.columns.duplicated(keep='first')]

    # Remove specific strings from the 'info' column
    df['info'] = df['info'].str.strip()  # Removes leading space
    strings_to_remove = ["Estimate Margin of Error", "Margin of Error", "Percent Margin of Error"]
    df = df[~df['info'].isin(strings_to_remove)]
    df = df.drop('info', axis=1)

    # Add year column
    df['year'] = '20{}'.format(name[1:])

    # Reorder columns
    leading_cols = ['tract', 'year']
    df = df[leading_cols + [col for col in df.columns if col not in leading_cols]]

    # Rename columns
    df.columns = ['tract', 'year', 'total_population', 'male', 'female', 'median_age',
                  'race_white', 'race_black', 'race_ai_alaskan', 'race_asian',
                  'race_hawaiian', 'race_other', 'race_hispanic']

    # Define the output filename - for exporting
    output_filename = data_folder + f"{name}_processed.csv"

    # Export the DataFrame to a CSV file
    df.to_csv(output_filename, index=False)

    print(f"{name} has been exported to {output_filename}")
```

3.2. Boston Bikes Infrastructure Data

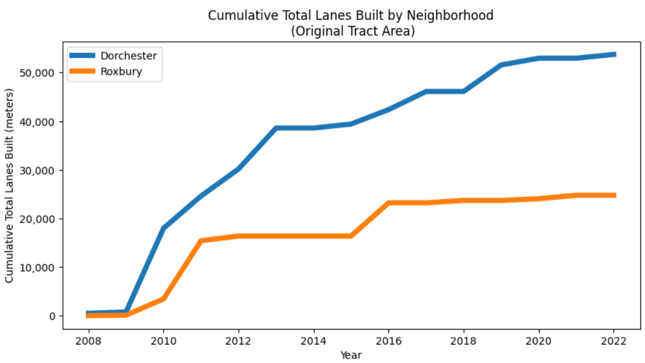
Our project's main goal is to examine the relationship between new infrastructure and socioeconomic changes within communities. After our first meeting with LivableStreets, we decided to focus our analysis efforts on the possible effects of new bike lanes. These bike lanes are meant to provide alternative transportation and commute options, while promoting leisure activities and health benefits for residents.

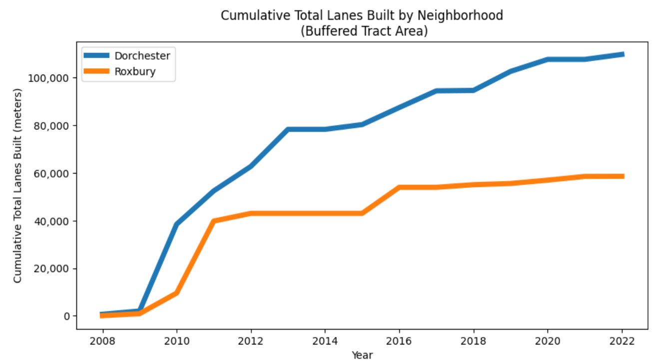
We used data on existing bike lanes in Boston from the [Boston Maps] project. The data, which was last updated January 2023, includes information on all existing bike lanes: geographic location, length, year of addition, and type (bike lanes have different forms and types: they can be simple signs on a car lane, indicating the shared use of the road; they can be physically built and painted lanes, with a vertical barrier between them and the road; or something in between). In our neighborhoods of interest - Roxbury and Dorchester, there are relatively few bike lanes that have some separation from the car lanes, and are mostly defined as simple bike lanes: a painted lane at the side of the road, not part of the car lane but without a physical/vertical separation. Since our data included the exact location of the bike lanes, we could match a given area - a census tract, in our case - and the bike lanes around it, at any point in time.

For determining which bike lanes are considered relevant for each tract, we decided to add a 800-feet "buffer" around the area of the tract, as defined in the Census data. The reason for increasing the size of the tracts is twofold: first, to account for the residents who live at the edges of the tract, and can benefit/use bike lanes that are placed just outside the tract boundaries. Second, bike lanes are often built on main streets, which also make natural boundaries between tracts. By having a small buffer we are avoiding an arbitrary assignment of a bike lane to only some of the tracts it borders. (This means that each bike lane can be counted more than once in our analysis, and most are - as can be seen in the following figure). This 800-ft buffer represents a conservative assumption of a 5-minute walking distance: we assume that having a bike lane within 5 minute walking distance might be relevant to one's decision whether to use the bike lane and to ride a bicycle on a regular basis.

Within our neighborhoods of interest, Roxbury and Dorchester, a total of ~78 kilometers of bike lanes were added between 2008 and 2023; with our "buffered" tract area calculation, this number more than doubles to ~ 170 kilometers (see Figure 3.3). Hereinafter we will use the "buffered" calculation as the total length of lanes attributed to each tract. As can be seen in the figures, Dorchester tracts had more than double the total length built than Roxbury tracts.

Figure 3.3 - Bike Lanes Built by Neighborhood (Original vs. Buffered Tracts)

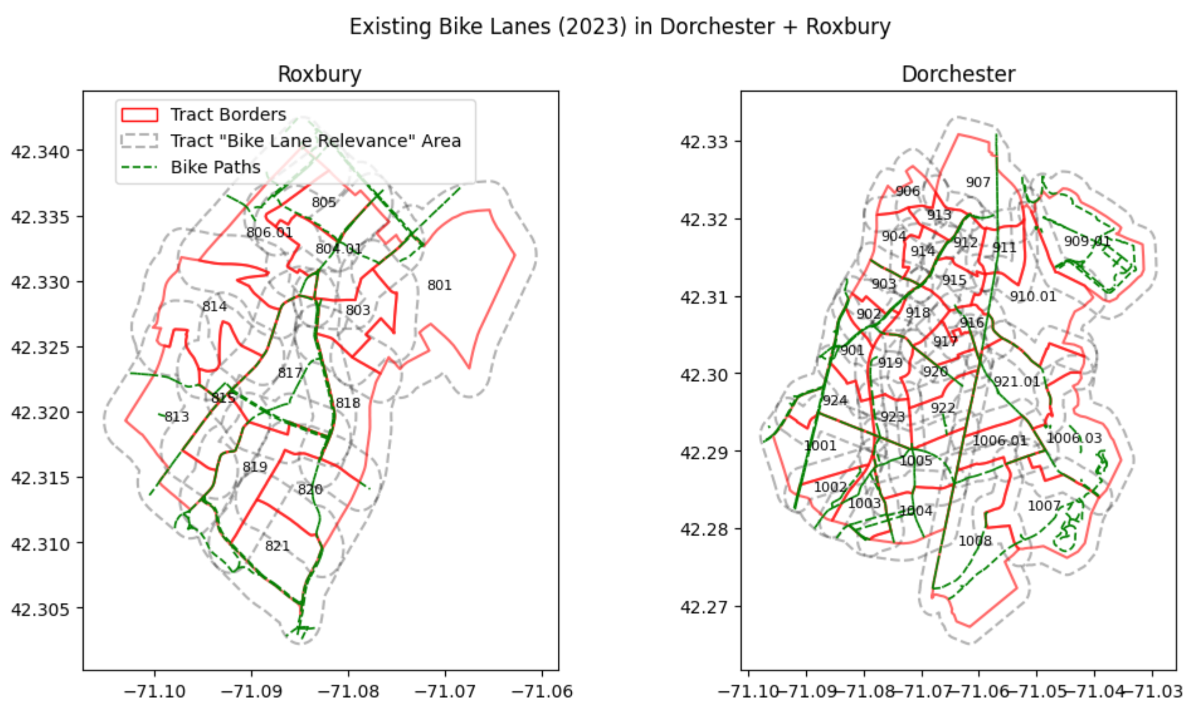




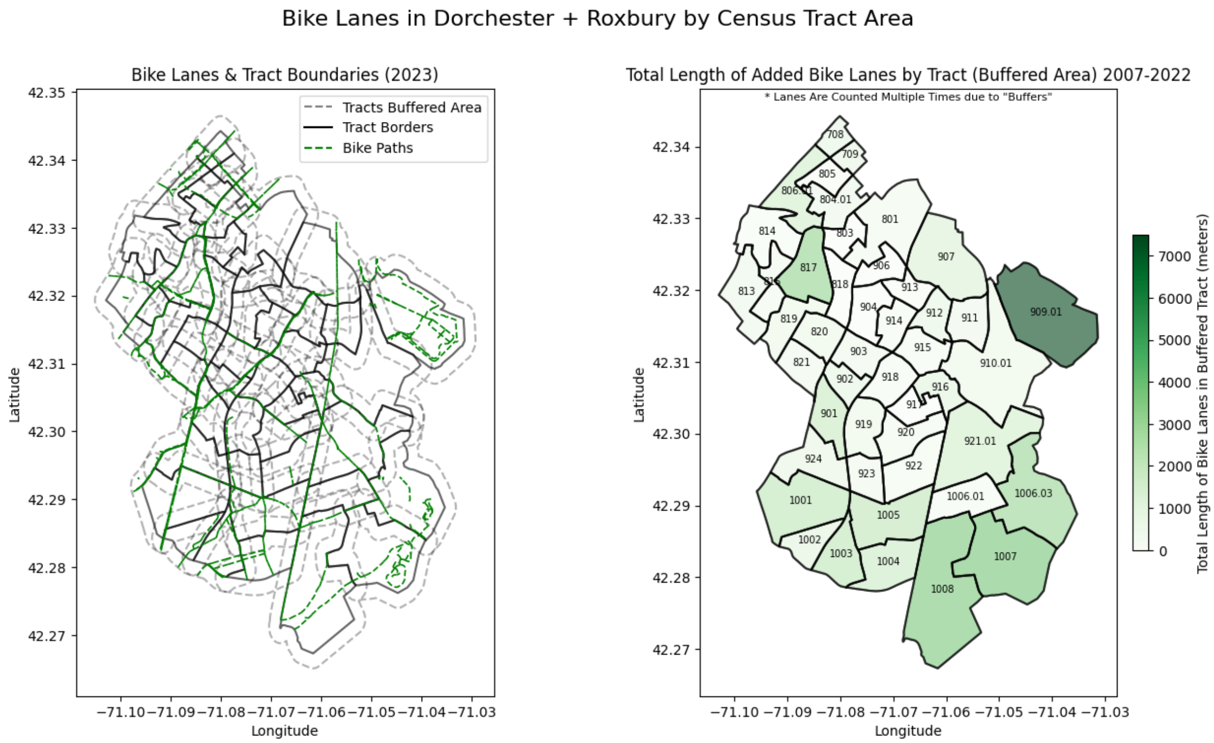
This results in a much more extensive network of bike lanes in Dorchester in 2023, as can be seen in the following map. We can also note that the bike lanes are spread across the neighborhoods on one hand, but there is a significant variation in each neighborhood, with some tracts having many/long lanes, and others less/shorter ones.

Figure 3.4 - Census Tract Maps of Bike Lanes

(1)



(2)



The variation between tracts and years allows us to explore the relationship between bike lanes and social and economic conditions: where are bike lanes being built? Can we point to differences between areas that have more bike lanes and areas with fewer? The correlation matrix in Table 3.1 summarizes the key relationships between bike lanes and social variables:

Table 3.1 - Correlation Between Bike Lane Development and Social Factors

	Median Income	Total Bike Lanes Length	% White Population	Median Age	% Home Owners
Median Income	1.00	-0.14	0.62	0.52	0.76
Total Bike Lanes Length	-0.14	1.00	-0.06	-0.11	-0.06
% White Population	0.62	-0.06	1.00	0.22	0.41
Median Age	0.52	-0.11	0.22	1.00	0.59
% Home Owners	0.76	-0.06	0.41	0.59	1.00

We see in Table 3.1 that different social and economic variables are highly correlated: areas with larger white populations also have higher income, median age, and proportion of homeowners. Bike lanes tend to be built in areas that have lower income, younger populations, fewer homeowners and smaller white population. This is a simple correlational observation, but we will look closer into this relationship in the regression analysis.

In Table 3.2, we look specifically within each neighborhood to see if these patterns are similar across different areas. We see similar patterns in both neighborhoods, but with some key differences: in both neighborhoods, new bike lanes are added to Census Tracts with lower income, lower median age, and lower proportions of white residents and homeowners. However, the magnitudes of these correlations are different: while in Dorchester the strongest negative relationship is between bike lanes and income, in Roxbury it is between bike lanes and white residents. This difference can be partially explained by the differences in the neighborhoods: Dorchester shows a much stronger correlation between income and race, or income and age, than Roxbury; additionally, in Roxbury the relationship between age & race is positive (higher proportion of white residents is associated with younger population) while Dorchester is the opposite. We also know from our preliminary analysis that Roxbury is more homogeneously non-white and lower income than Dorchester whose distributions for these same features are broad.

Table 3.2 - Neighborhood-Specific Feature Correlations

Neighborhood		Median Income	Total Bike Lanes Length	% White Population	Median Age	% Home Owners
Dorchester	**Median Income**	1.00	-0.17	0.70	0.63	0.71
	Total Bike Lanes Length	-0.17	1.00	-0.04	-0.12	-0.06
	% White Population	0.70	-0.04	1.00	0.44	0.50
	Median Age	0.63	-0.12	0.44	1.00	0.72
	% Home Owners	0.71	-0.06	0.50	0.72	1.00
Roxbury	**Median Income**	1.00	-0.06	-0.07	0.37	0.71
	Total Bike Lanes Length	-0.06	1.00	-0.12	-0.09	-0.02
	% White Population	-0.07	-0.12	1.00	-0.56	-0.18
	Median Age	0.37	-0.09	-0.56	1.00	0.40
	% Home Owners	0.71	-0.02	-0.18	0.40	1.00

4. Exploratory Analysis and Communities of Interest

The communities of Dorchester, Roxbury, and Jamaica Plain in the Boston Metro Area serve as focal points for our exploratory analysis at the intersection of transportation infrastructure and displacement.

These neighborhoods exhibit unique socioeconomic dynamics, making them crucial areas for investigation. Economic and social indicators such as race, property value, home ownership, and income will be central to our analysis. Dorchester, known for its cultural diversity, faces ongoing challenges related to economic disparities. Roxbury, with its rich history, has experienced shifts in property values and homeownership patterns. Jamaica Plain, characterized by a mix of socio-economic backgrounds, presents an opportunity to explore how transportation projects influence different income groups. By examining these economic indicators, we aim to unravel the intricate relationship between transportation initiatives, such as bus lanes and bike infrastructure, and potential displacement, providing valuable insights for targeted policy interventions and community-focused development strategies.

4.1. Race

In Figure 4.1 each point represents the population percentage of each race in one Census Tract in a neighborhood. For example, in the 2010 plot, Census Tract 1101.03 in Jamaica Plain has 6 dots on the left graph, only one of which is assigned to the 'white' category. The other 5 correspond to (1) Black, (2) American Indian and Native Alaskan, (3) Asian, (4) Hawaiian, and (5) other races.

Figure 4.1 - Race and Neighborhood Analysis

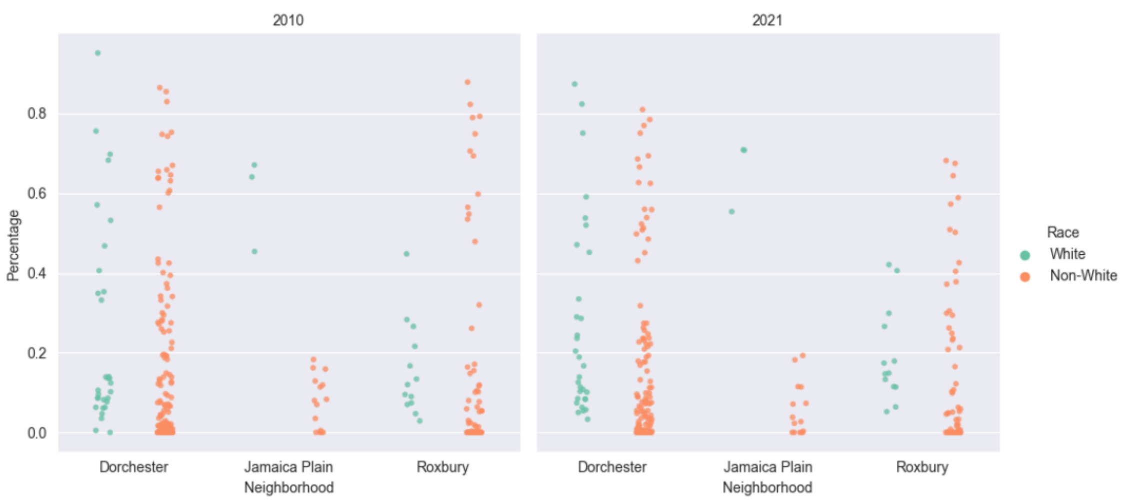
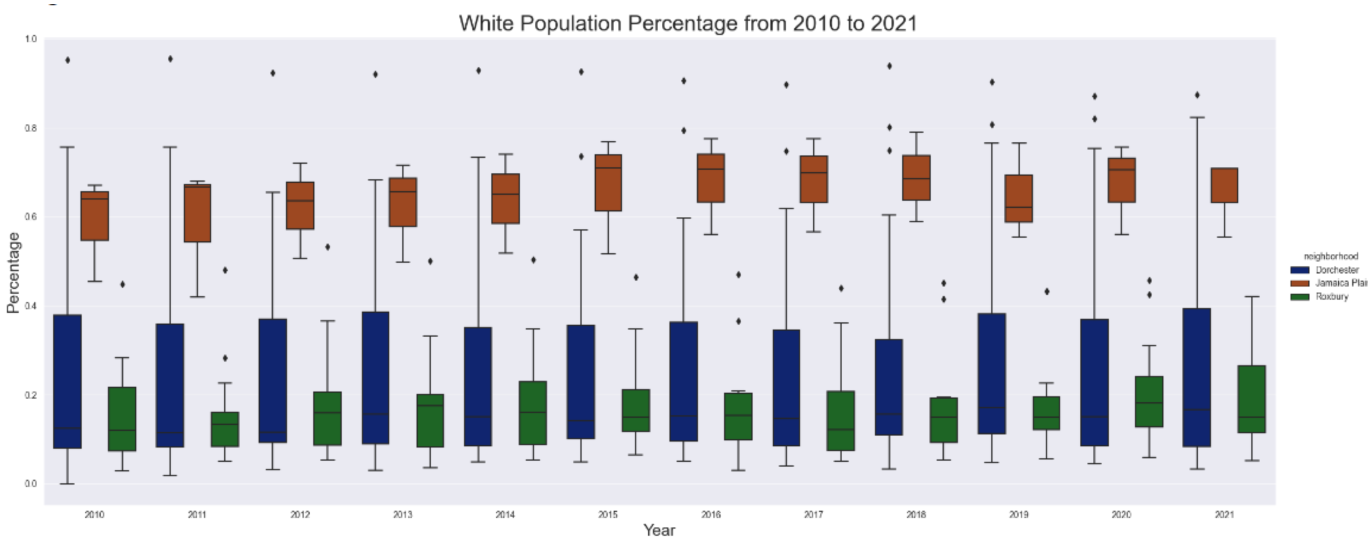


Figure 4.1 shows that Jamaica Plain has the largest disparity between white and non-white populations. On the other hand, Roxbury's Census Tracts are typically associated with much smaller white populations. Between 2010 and 2021, the proportion of non-white residents decreased across all Census Tracts.

Figure 4.2 - Composition of Race Demographic for Each Neighborhood



The above graph makes clear the high median population of white residents in Census Tracts 1101.03, 1201.04, 1202.01 of Jamaica Plain, which surround the Forest Hills neighborhood, compared to all of Roxbury and Dorchester. Also, while Roxbury's Census Tracts have consistently low white populations, Dorchester displays strong right-skewness with regard to white residency. This means that while the median of white populations is low, there are areas of Dorchester with large to almost totally white populations.

4.2. Home Ownership

Figure 4.3 - Homeownership and Neighborhood Analysis

(1)

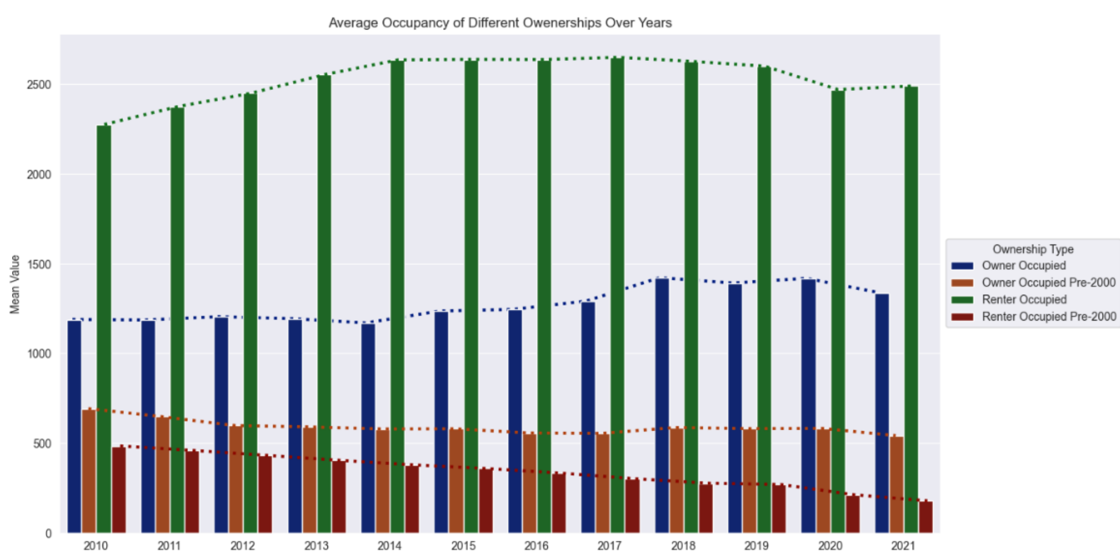
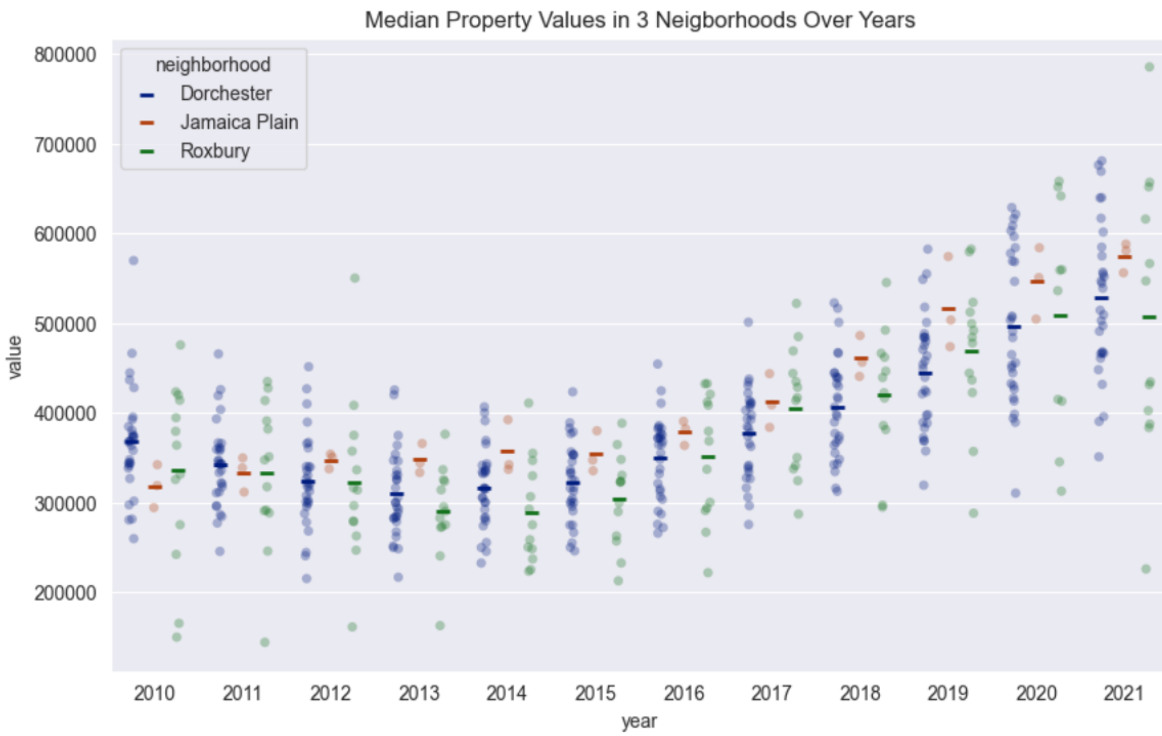


Figure 4.3 demonstrates that these neighborhoods are predominantly composed of renters. Of those, the renters who moved into their neighborhoods prior to 2000 leave at a higher rate than owners who moved in during the same era. While the measurement of residents who moved in pre-2000 cannot increase year-over-year, its rate of decrease can inform us how quickly older populations are leaving their neighborhoods.

4.3. Property Value

Figure 4.4 - Property Value and Neighborhood Analysis



In Figure 4.4, we see that the 2010 median property value in all three neighborhoods was around $350,000. After that, property value in Jamaica Plain went on a steadily increasing trend, while Roxbury's and Dorchester's median property values decreased for a few years before increasing around 2015.

Figure 4.5 - Mortgage and Rent Analysis in Aggregate

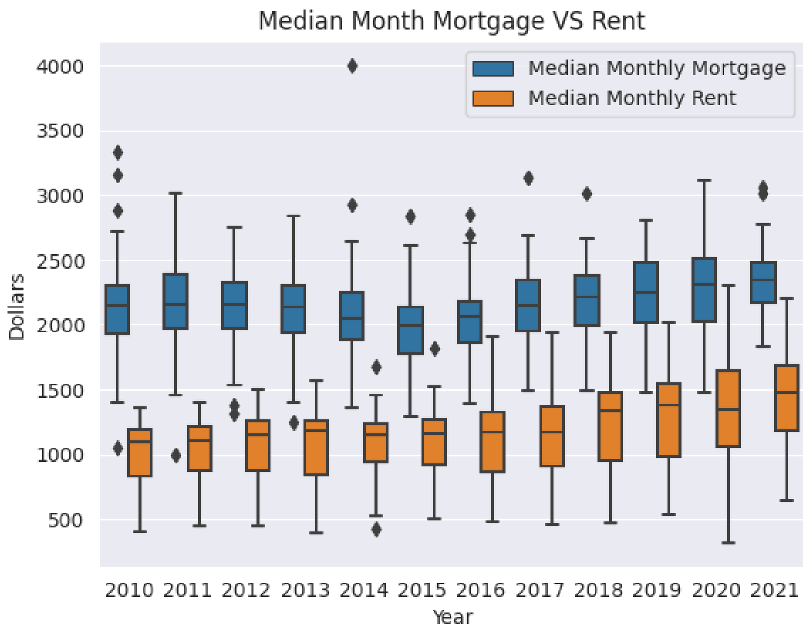


Figure 4.5 is the comparison between median monthly mortgage and monthly rent without differentiating different neighborhoods. From 2010 to 2021, median monthly rent has been smoothly increasing in general, while median monthly mortgage slightly experienced a dip from 2013 to 2017.

Figure 4.6 - Mortgage and Rent Payments by Neighborhood

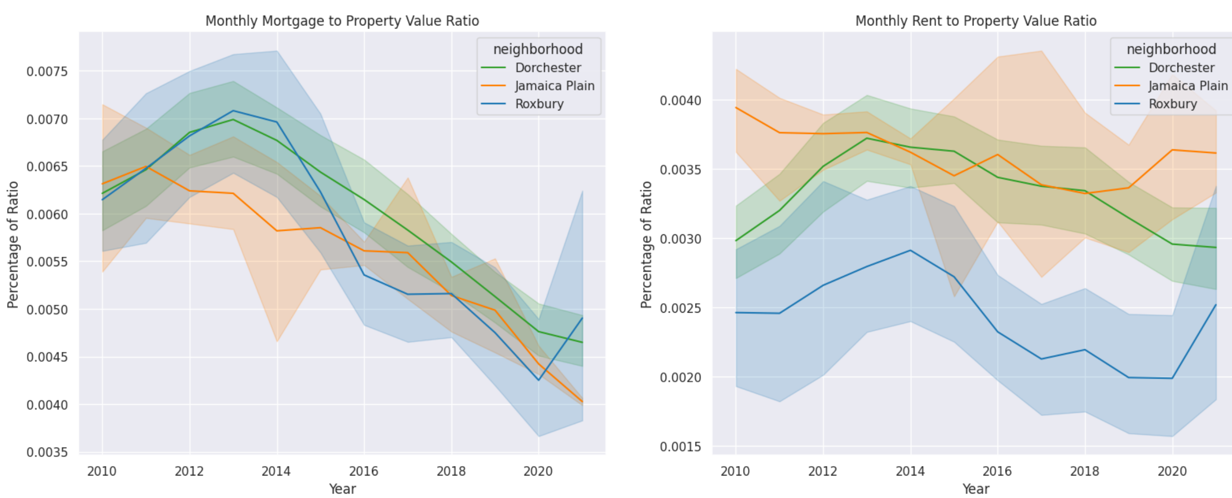


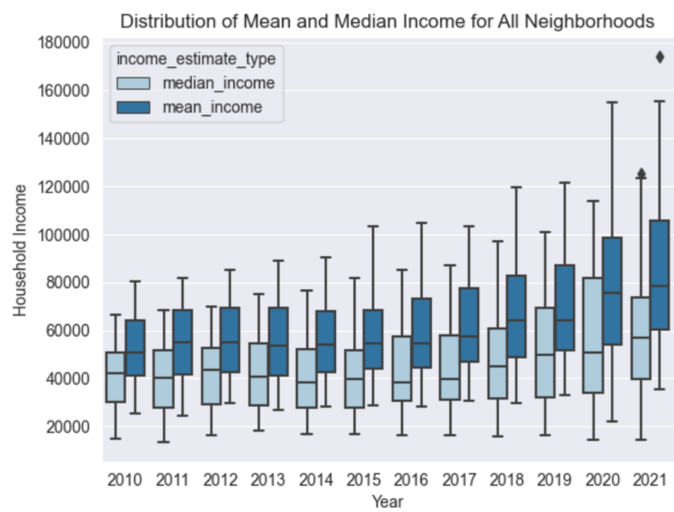
Figure 4.6 shows the trend for both monthly mortgage and monthly rent in different neighborhoods. All three neighborhoods show similar behavior on the general trend for monthly mortgage to property value ratio, which has decreased relative to the ratio in 2010. However, Dorchester and Roxbury exhibited an upward trend from 2010 to approximately 2013.

While all three neighborhoods' monthly mortgage to property value ratio do not differ much, it is not the same case for their monthly rent to property value ratio; Roxbury has been associated with a comparatively lower monthly rent to property value ratio during the entire 2010-2021 time period.

4.4. Income

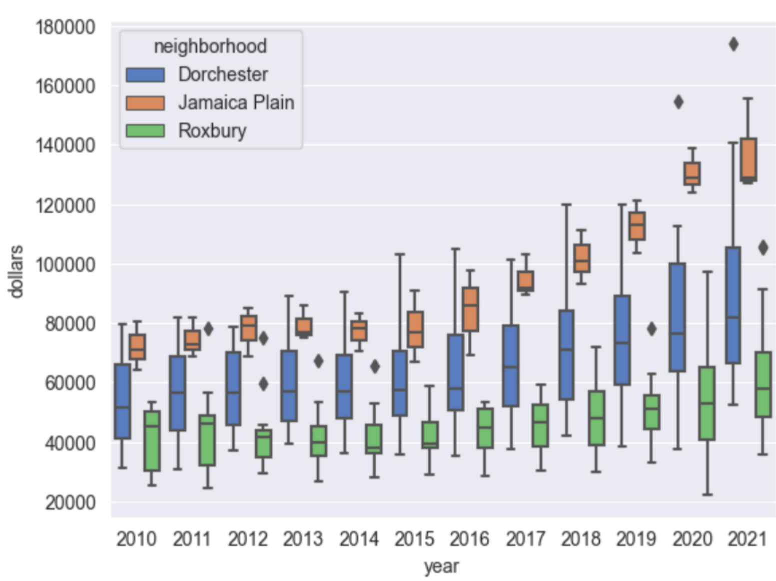
The degree of separation between mean and median income in Figure 4.7 can be used to assess the degree of income inequality. If a minority upper class experiences an increase in household income, mean income would increase more drastically than median income, which better describes the middle class.

Figure 4.7 - Income Analysis in Aggregate



This is precisely the effect we observe above. Each year, the aggregate median household income for Roxbury, Dorchester, and Jamaica Plain moves further away from the median.

Figure 4.8 - Income Analysis by Neighborhood

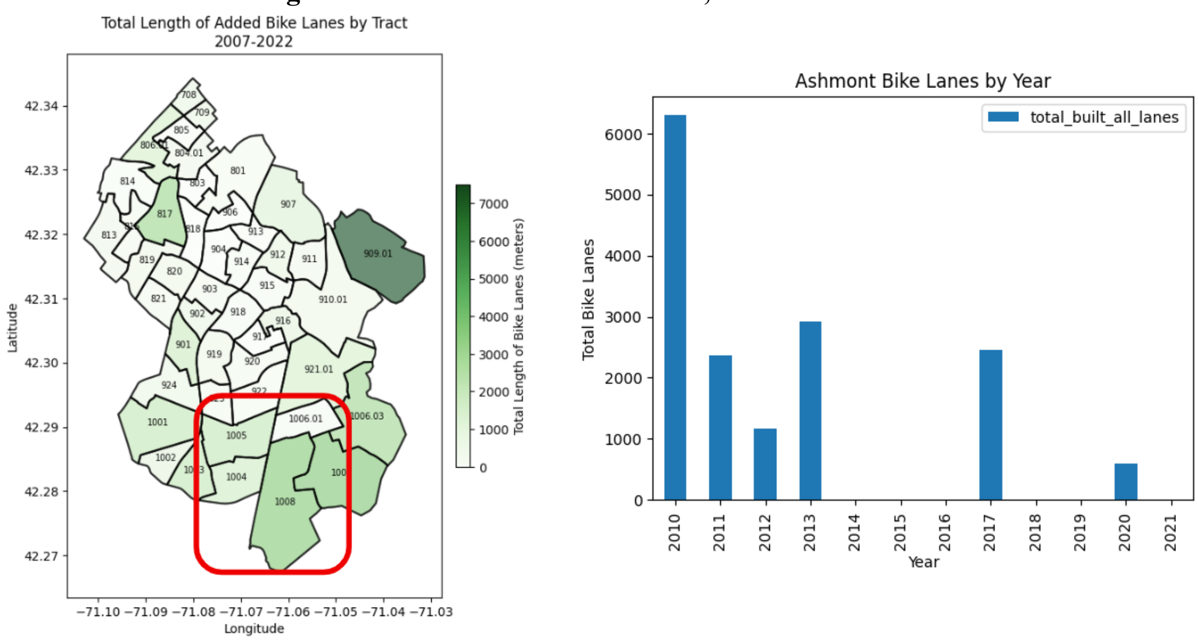


When we separate the neighborhoods in Figure 4.8, it is clear to see that the predominantly white neighborhoods in Jamaica Plain's increase in mean household income accounts for the higher degree of spread between mean and median income. Year over year, Roxbury's median average income increases only slightly, and Dorchester's still less than Jamaica Plain's. It is worthy to note here that, as with race, Dorchester has the most diverse set of average incomes.

4.5. Ashmont-DOT Greenway Analysis

The Ashmont neighborhood of Dorchester is part of an ongoing project called the Dorchester (DOT) Greenway, which seeks to install a pedestrian and bike-friend greenway along the MBTA Red Line tunnel cap (LivableStreets Alliance). LivableStreets says that this project will fill a critical missing link in the existing greenway network, creating a connection between Franklin Park, the Neponset River Greenway, and the Harborwalk, but the project has been met with the same resistance they've come to know. They requested we look at Ashmont in comparison with the rest of Dorchester, particularly with regard to the likelihood of change alongside bike lane development.

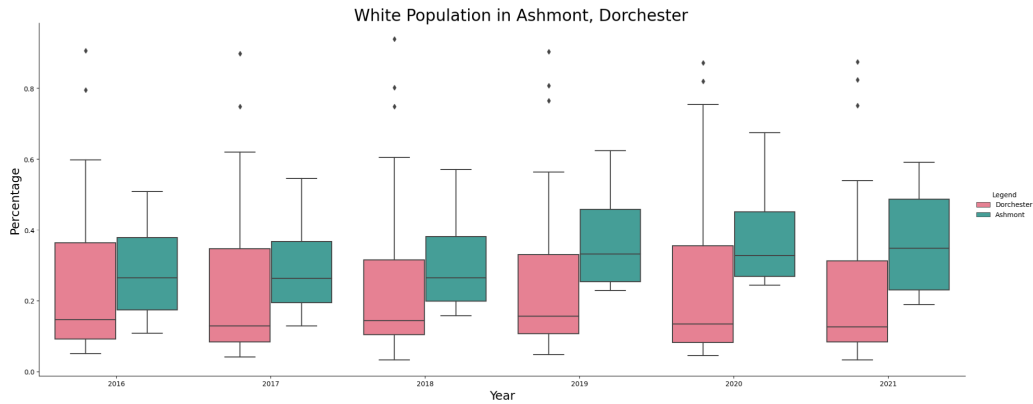
Figure 4.9 - Bike Lanes in Ashmont, Dorchester



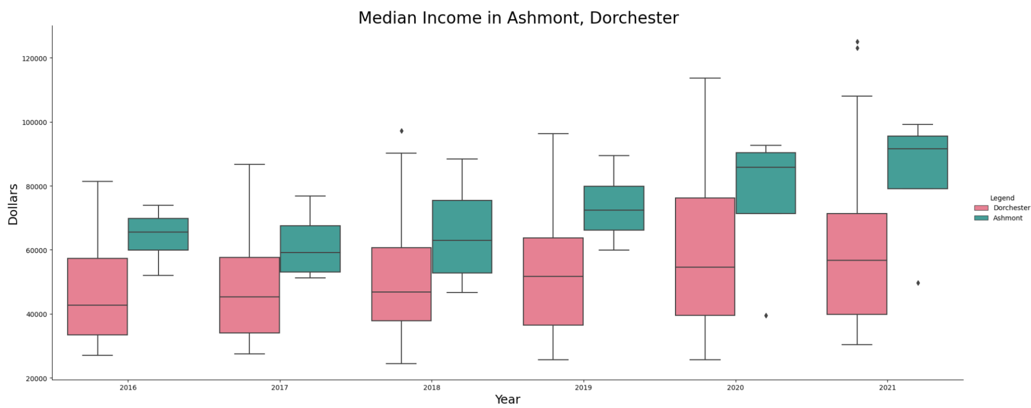
Ashmont comprises parts of Census Tracts 1004, 1005, 1006.01, and 1008, as shown in the left panel of Figure 4.9. These have seen relatively high bike lane development since 2010 although progress has slowed significantly in recent years, as documented in the right panel.

Figure 4.10. Demographic Assessment of Ashmont

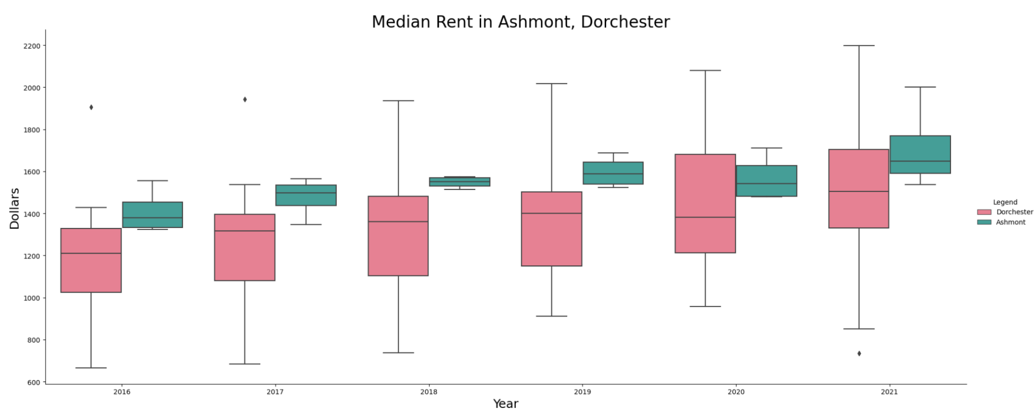
(1)



(2)



(3)



(4)

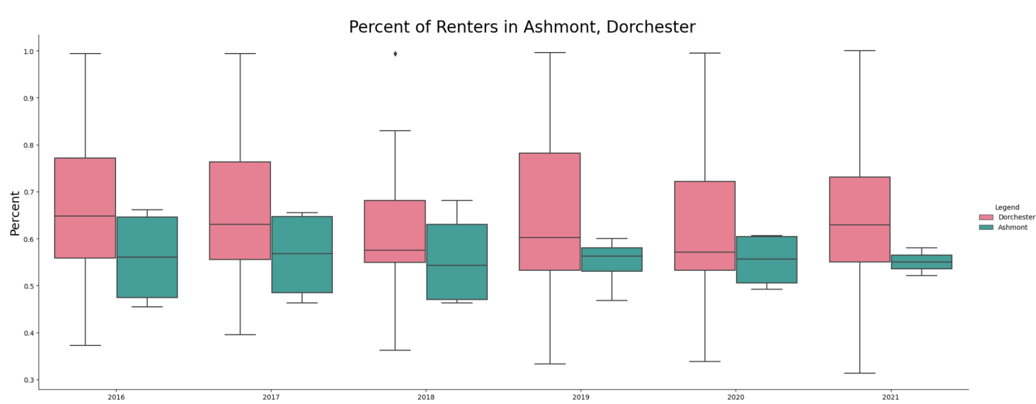
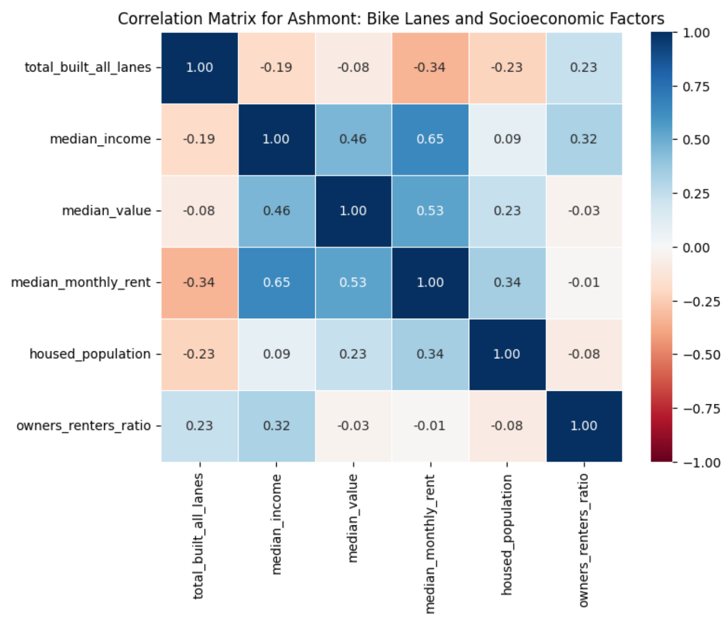


Figure 4.10 compares Ashmont's key factors for displacement (shown in green) with those of the rest of Dorchester excluding Census Tracts 1004, 1005, 1006.01, and 1008 (shown in red). We find that (1) Ashmont has higher median white population with its median measuring above Dorchester's 75th percentile since 2019; (2) Ashmont's median income has exceeded Dorchester's 75th percentile for the last 6 years; (3) median rent in Ashmont has been approaching that of Dorchester's, signaling a decline in Ashmont's rate of rent increase; and (4) Ashmont has consistently fewer renters, and more owners, than the rest of Dorchester.

When we take a look at these factors and their relationship with bike lane development, we find the correlation matrix shown in Figure 4.11. This matrix computes the linear correlation between the variable "total_built_all_lanes," a feature which computes the per year bike lane development for the Census Tracts in Ashmont, and other selected social factors for the same neighborhood.

4.11 - Ashmont Social Factor-Bike Lane Development Correlation Matrix



We observe a moderate, negative correlation between bike lane development and monthly rent. This indicates that where bike lanes are built in Ashmont, there is generally *lower* rent. While this is not a statement of causality, it could illustrate that bike lanes are built in areas with lower property values. Beyond this, there is a low, negative correlation between bike lane development and median income. Following the same logic, this may show that bike lanes are developed in *lower* income areas of Ashmont. Both of these observations align with LivableStreets' mission to provide equitable transportation access by dismantling barriers to even the least privileged populations.

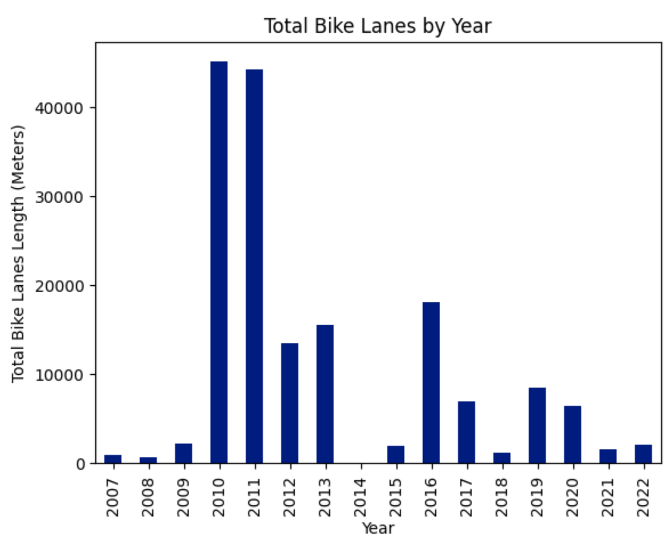
In the context of LivableStreets' mission, this analysis supports continuing the DOT Greenway initiative. There is no positive correlation between bike lane development and any factor of displacement that we identified. While correlation does not imply causation, we aim to uncover if a causal relationship exists between these variables in the next section.

5. Aggregate Effect of Bike Infrastructure on Communities

Our project's null hypothesis can be summarized as the following: **socio-economic variables in residential areas are not impacted by the addition of new bike lanes**. This means that, on average, we should see similar trends in areas which experience new bike lanes and in areas which don't. This *does not* mean that the absolute values of these variables should be similar between these area groups; moreover, we have already seen that bike lanes are being built, on average, in areas with lower income, less white population, lower house values, etc. In this part we wish to examine the possibility of a causal relationship: is the addition of bike lanes associated with a change in key variables in the following years? While establishing a cause and effect relationship is a significant challenge, here we are assessing a key part of it: using a time-series regression, we test whether new lanes are followed by a socio-economic change.

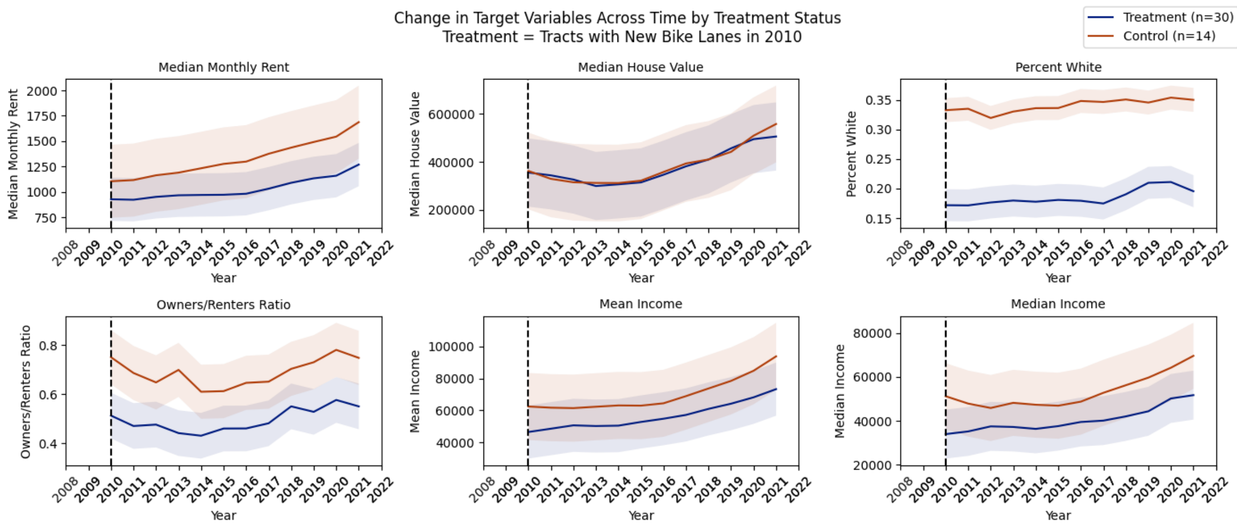
New bike lanes were being added throughout the years in our data. This makes it challenging to identify the effect of an individual "treatment"; we choose to use two years with significant total length of new lanes: 2010 and 2016. As can be seen in Figure 5.1, 2016 had significantly more new bike lanes compared to the previous and following couple of years, and thus can arguably be used as a "treatment" year. It's important to note that, as mentioned previously, the bike lane length figures include "double-counting" where each lane is counted once for each Census Tract and the "buffer" it intersects.

Figure 5.1 - Bike Lane Construction by Year



Next, we will discuss the interpretation of each year's model results. For 2010, the regression is looking only "forward", since our census data starts at 2010; it compares tracts in which there were new bike lanes in 2010 (treatment group) with those in which there wasn't (control). As the following plots demonstrate, while there are notable differences in the absolute levels of the socio-economic target variables, the trends are quite similar across time.

Figure 5.2 - Difference-in-Difference Regression Analysis for 2010



We ran a linear regression with an interaction term between the treatment dummy variable and the year to estimate the effect of the treatment over time. Following our discussion with LivableStreets, we limited the test to 1-5 years after the building of bike lanes, i.e., years 2011-2016.

For each target variable (Median Income, Median House Value, Median Monthly Rent, Percentage of White Population, and House Owners/Renters Ratio), the regression is estimated in the following form:

$$Target = \beta_0 + \beta_1 * Treatment + \beta_2 * Year + \beta_3 * Treatment * Year + \beta_4 * Tract + \varepsilon$$

Since the data is organized in (tract, year) rows, the variables are representing an amount in a single tract, in a single year.

The following is an example of a complete (truncated) model result summary and a summary table for the R-squared, interest variable coefficient (Treatment * Year) and p-value for each of the target variables.

Figure 5.3 - Regression Summary for 2010

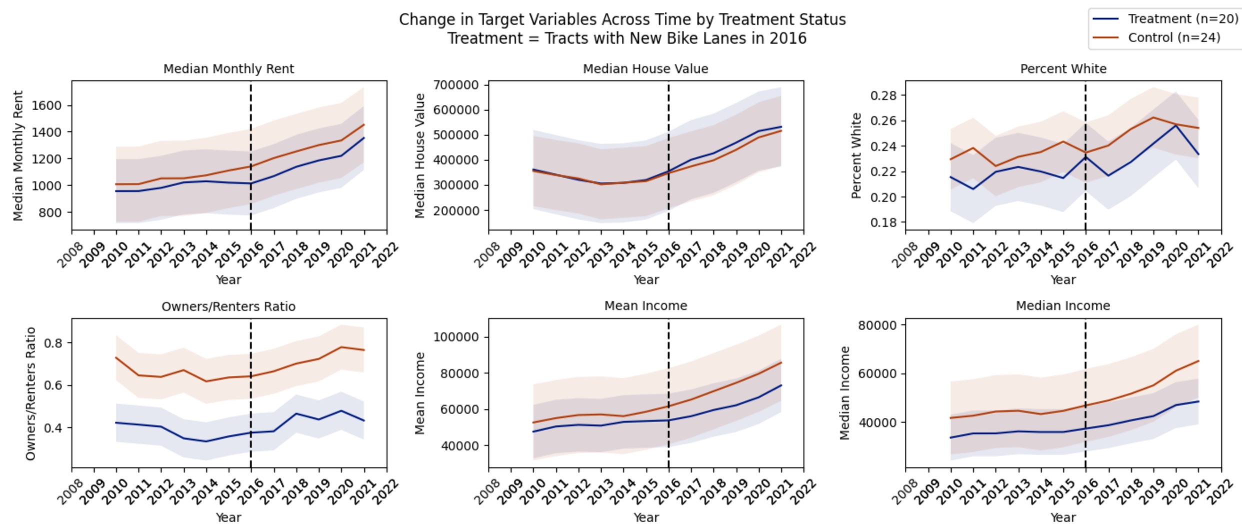
```
                            OLS Regression Results
==============================================================================
Dep. Variable:          median_income   R-squared:                       0.922
Model:                            OLS   Adj. R-squared:                  0.909
Method:                 Least Squares   F-statistic:                     69.15
Date:                Tue, 28 Nov 2023   Prob (F-statistic):          5.90e-121
Time:                        00:22:49   Log-Likelihood:                -3032.1
No. Observations:                 308   AIC:                             6156.
Df Residuals:                     262   BIC:                             6328.
Df Model:                          45
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
Intercept           -1.248e+06   4.41e+05     -2.830      0.005   -2.12e+06    -3.8e+05
treatment           -1.031e+05   5.46e+05     -0.189      0.850   -1.18e+06    9.71e+05
year                  683.9668    249.850      2.738      0.007     191.998    1175.936
treatment:year         -0.2109    302.583     -0.001      0.999    -596.015     595.593
==============================================================================
Omnibus:                        8.258   Durbin-Watson:                   1.442
Prob(Omnibus):                  0.016   Jarque-Bera (JB):                9.578
Skew:                           0.270   Prob(JB):                      0.00832
Kurtosis:                       3.674   Cond. No.                     1.00e+19
==============================================================================
```

Target Variable	R2	Coefficient	P-value
Median Income	0.922	-0.211	0.999
Median House Value	0.506	-4245.183	0.143
Median Monthly Rent	0.908	-26.364	0.000
% White Population	0.983	-0.003	0.101
Home Owners/Renters Ratio	0.919	0.008	0.261

The regression results show a non-significant coefficient for our interest variables: median income; median house value; white population proportion; and owners/renters ratio. In all of these, the data supports the null hypothesis, that tracts with newly added bike lanes did not experience significantly different trends, compared to tracts without them. One regression results do show a statistically significant coefficient for the interaction term: the monthly rent regression. The direction of this coefficient, however, is negative: tracts in the treatment group experienced *slower* increase in rent prices, compared to the control group - contrary to the concern of bike lanes driving up prices. This can also be seen in the plot shown above. We should note, however, that the size of the coefficient is quite small, and shows an average of ~$26 lower median rent in the treatment group tracts. To conclude, the 2010 bike lanes do not provide evidence for the concern of displacement.

For 2016, the regression is looking both backward and forward, and is implementing a difference-in-difference, or “diff-in-diff”: comparing the trends from before and after 2016, marking 2016 as the “treatment year”. The regression results for 2016 are focusing on the interaction term *Treatment * Post-2016 * Year* as the variable of interest: if this variable is statistically significant, it means we cannot reject the null hypothesis - that there is a difference in the trends between the treatment and control groups after the addition of the bike lanes.

Figure 5.4 - Difference-in-Difference Regression Analysis for 2016



Here, we ran a linear regression with an interaction term between the treatment dummy variable, year, and the post-treatment variable to estimate the effect of the treatment over time (*Treatment * Post-Treatment * Year*). This is different from the 2010 regression in that we are considering the difference between *before* and *after* the treatment, which is in 2016. So we are estimating the difference in trends in our target variables, before vs after the treatment. We limited the scope of the data to include a 5-year interval for both times: 2011-2015 for before, and 2017-2021 for after.

For each target variable (Median Income, Median House Value, Median Monthly Rent, Percentage of White Population, and House Owners/Renters Ratio), the regression is estimated in the following form:

$$\begin{aligned} Target = {} & \beta 0 + \beta 1 * Treatment + \beta 2 * Post\text{-}Treatment + \beta 3 * Year + \\ & \beta 4 * Treatment * Post\text{-}Treatment + \beta 5 * Post\text{-}Treatment * Year + \\ & \beta 6 * Treatment * Year + \beta 7 * Treatment * Post\text{-}Treatment * Year + \beta 8 * Tract + \varepsilon \end{aligned}$$

Here, our coefficient of interest is Treatment β7 (*Treatment * Post-Treatment * Year*). If this coefficient is statistically significant, it indicates that on average, tracts with new lanes experienced a shift in one or more of the target socio-economic variables post 2016 - i.e., after new bike lanes were added in the area.

Following are an example for a complete (truncated) model result summary and a summary table for the R-squared, interest variable coefficient and p-value for each of the target variables.

Figure 5.5 - Regression Summary for 2016

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     owners_renters_ratio   R-squared:                       0.896
Model:                              OLS   Adj. R-squared:                  0.884
Method:                   Least Squares   F-statistic:                     75.79
Date:                  Tue, 28 Nov 2023   Prob (F-statistic):          4.10e-181
Time:                          00:45:22   Log-Likelihood:                 318.62
No. Observations:                   480   AIC:                            -537.2
Df Residuals:                       430   BIC:                            -328.6
Df Model:                            49
Covariance Type:              nonrobust
==================================================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------
Intercept                         5.0864     12.933      0.393      0.694     -20.333      30.506
treatment                        14.2695     18.269      0.781      0.435     -21.638      50.177
post_treatment                  -61.8683     21.516     -2.875      0.004    -104.158     -19.579
treatment:post_treatment         10.5585     31.982      0.330      0.741     -52.302      73.419
year                             -0.0023      0.006     -0.364      0.716      -0.015       0.010
treatment:year                   -0.0074      0.010     -0.781      0.435      -0.026       0.011
post_treatment:year               0.0307      0.011      2.877      0.004       0.010       0.052
treatment:post_treatment:year    -0.0052      0.016     -0.329      0.742      -0.036       0.026
==============================================================================
Omnibus:                       28.278   Durbin-Watson:                   1.168
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               47.455
Skew:                           0.408   Prob(JB):                     4.96e-11
Kurtosis:                       4.307   Cond. No.                     6.64e+18
==============================================================================
```

Target Variable	R2	Coefficient	P-value
Median Income	**0.89**	**-1449.12**	**0.09**
Median House Value	**0.76**	**-4921.33**	**0.44**
Median Monthly Rent	**0.87**	**22.07**	**0.16**
% White Population	**0.97**	**0.00**	**0.78**
Home Owners/Renters Ratio	**0.90**	**-0.01**	**0.74**

Similarly to the 2010 results, the regression with 2016 data does not support evidence for rejecting the null hypothesis. We see that along with high R-squared values, the interaction term of interest is not-significant at the 95% level in any of the models. We have one model with a coefficient that is significant at the 10% level - median income. Again, the sign of this coefficient is the opposite of what we would expect, if we believed bike lanes were a cause of displacement: on average, tracts in the treatment group saw a lower increase in income (both median and mean, s can be seen in the plots). Thus we can conclude that none of the models we tested point in the direction of the causality some residents might fear: tracts in which bike lanes are built might have different characteristics than tracts in which they are not, but the trends of these characteristics across time are similar, and at least not different at a statistically significant level.

6. Casey Arborway Case Study

The purpose of including Forest Hills in our analysis is to offer an alternative form of transportation infrastructure development with which to compare bike lanes. In 2019, construction was completed on the Casey Arborway, a new greenway which replaced the defunct Casey Overpass traffic bridge through Forest Hills. Originally built in 1955, the overpass had fallen into such disrepair that two of the four lanes had to be shut in the first decade of the new millennium. LivableStreets advocated for the Arborway beginning in 2011, specifically the at-grade option, which would replace the overpass with a street-level traffic pattern. Ultimately, this was the proposal that the Department of Transportation approved.

To assess changes in the factors of displacement before and after completion of the Casey Arborway, we plotted the distribution of each for the Census tracts immediately surrounding the project.

Figure 6.1 - Income Over Time in Forest Hills

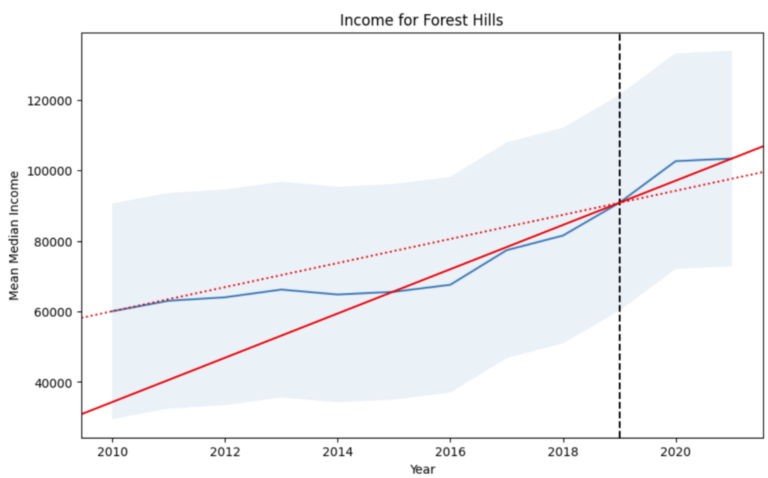


Figure 6.1 shows the distribution of income in Forest Hills shadowed by its confidence interval. The dashed red slope line indicated the trend from 2010-2019, whereas the solid red slope line projects the slope of data from 2020-2021.

Figure 6.2 - Race Over Time in Forest Hills

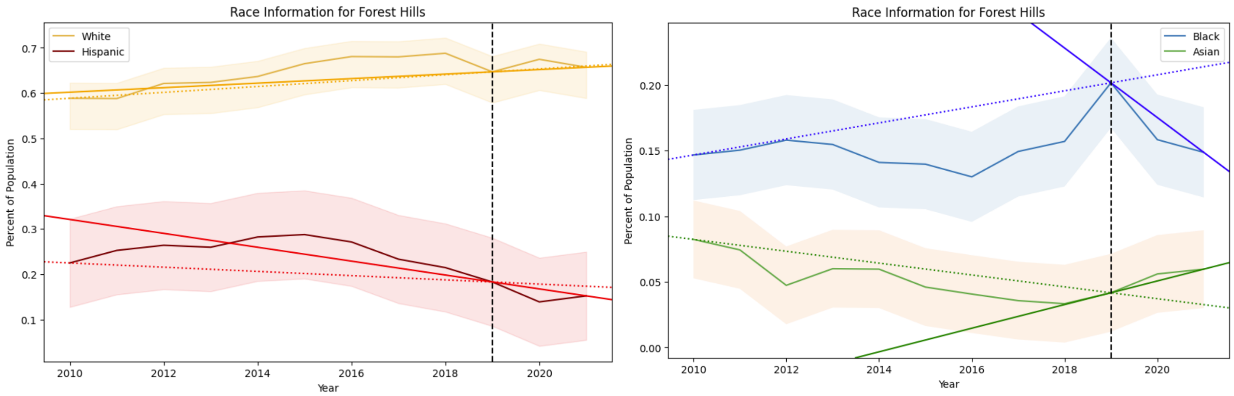


Figure 6.2 plots the racial composition of Forest Hills with 2019 as a breakpoint. On the left, it is clear to see that the trends for White and Hispanic residents are nearly the same when considering the period before and after the completion of the Casey Arborway. However, 2020 marked a steep decline in Black residents in Forest Hills while the Asian population experienced a gradual increase contrary to previous years. Here again, dashed lines represent the trends of 2010-2019 data and solid lines represent that of 2020 and 2021.

Figure 6.3 - Home Ownership Over Time in Forest Hills

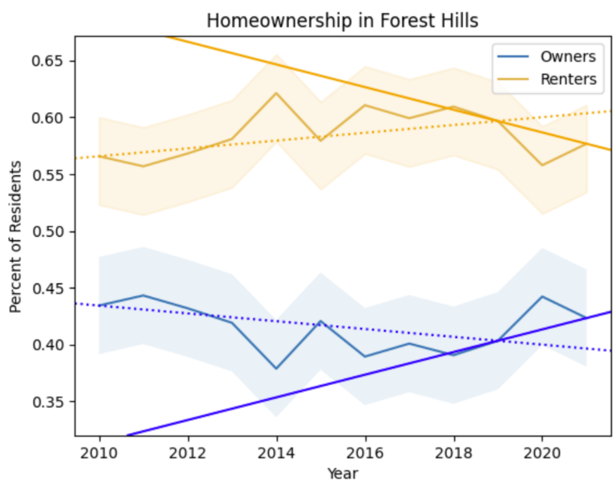


Figure 6.3 highlights a decline in renters post-2019 and a commensurate increase in homeowners. This indicates that, unlike the situation in South Mattapan, the construction of the Casey Arborway likely did not trigger development of apartment complexes en masse. However, a positive change in percentage of home ownership could indicate a change in residents' desire to live in the area for a longer term and an increase in their financial status.

Figure 6.4 - Property Value Over Time in Forest Hills

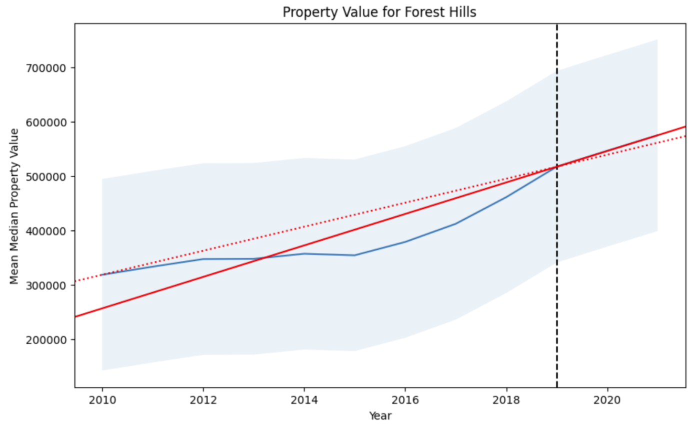


Figure 6.4 demonstrates little change in the trend of property values pre- and post-Casey Arborway construction.

Observing changing trends is one thing, but inferring causality requires regression analysis that shows statistical significance in the interaction between two variables. As we did in the previous section for Roxbury and Dorchester, we conducted a difference-in-difference analysis of each of the factors contributing to displacement for Forest Hills. The treatment group are the ACS 5-year estimates post-2019, and the control are the estimates for the same Census tracts two years prior to the Casey Arborway's completion.

With our limited data, we cannot conclude that treatment caused an effect in any of the variables: income, property value, monthly rent, or proportion of White residents. The difference-in-difference analysis yielded p-values of 0.11, 0.49, 0.64, and 0.79 for treatment on each of these variables, respectively. These are all above the threshold of 0.05 to reject the null hypothesis, therefore we cannot make claims of causality. Thus, we can only point to some partial evidence of a "positive" change in socioeconomic factors in Forest Hills following 2019, which could potentially force residents to other areas.

7. Conclusions

Considering that our analysis shows that we cannot reject the null hypothesis that bike lane development in Roxbury and Dorchester has no effect on income, race, property value, or homeownership, it does not stand to reason that bike lane development should be restricted in these areas when it has a tangible benefit to the community. Similarly, greenways like the Casey Arborway in Jamaica Plain, which connected the Forest Hills neighborhood to the Emerald Necklace, cannot be attributed to a causal effect on these social factors.

Another important finding is the correlation between new bike lanes and lower socioeconomic areas; given that bike lanes are built, among other reasons, to improve quality of life and provide alternative transportation options for residents, this finding should support existing policies, at least in the areas our study looked into. Combined with other considerations, economic and demographic factors should be taken into account when building new infrastructure: who are the residents in this area? Can they benefit from this new bike lane? It can be argued that areas with fewer private vehicles and higher reliance on public transport can benefit bike lanes significantly. Thus, our analysis can provide support for the allocation policies of new bike lanes in Roxbury and Dorchester in previous years.

While the statistical power of our analysis is low due to the localized nature of our study, we are confident in these results. It is likely that, as articles mentioned herein have stated, a combination of social and economic factors lead to the displacement rather than a single one. Nevertheless, this area of study is important to the many thousands of residents who worry about clearing their next housing payment in neighborhoods where they have established roots for generations. We would like to see this study expanded to the whole of Boston and extended in the next decade beyond the COVID-19 pandemic.

These conclusions match the findings of Ferenchak and Marshall (2021), who looked at more than 11,000 bike facility miles over 10 years in 29 cities and found that no/weak correlations exist between bike lanes and displacement.

8. Limitations

It is essential to acknowledge that our research endeavors are accompanied by notable challenges, demanding a measured awareness of their impact on the scope and interpretation of our findings.

Scope Constraints: Adhering to customer specifications and time constraints, our primary analysis is confined to three specific neighborhoods in Boston—Roxbury, Dorchester, and Jamaica Plain. While these areas may serve as reasonable proxies for socioeconomic characteristics in other Boston neighborhoods, it is important to acknowledge that the statistical robustness of our analysis is limited due to the inclusion of only three neighborhoods. Furthermore, our analysis for Jamaica Plain considers only 3 Census Tracts out of 14 as opposed to Roxbury and Dorchester where we look at the neighborhood in its entirety.

Generalization Limitations: Beyond the challenge of a limited geographical focus, our ability to generalize findings to broader regions is constrained. The utilization of bike lane data specific to Boston introduces variations in quality, maintenance, and other factors that may significantly differ from those in other locations. Disparities in road infrastructure, traffic conditions, and alternative transportation options across cities further emphasize the need to restrict the applicability of our analysis solely to the Boston area. Additionally, due to the structuring of our data, we cannot present particular effects as they relate to specific populations (e.g., the effect of bike lane development on income of Black populations in Roxbury).

Identification Challenges: The core objective involves identifying a causal relationship between new bike lane development and displacement. However, both the introduction of bike lanes and changes in socioeconomic variables occur simultaneously over our data's time range, posing a challenge in establishing causality. Although we used difference-in-difference methods to address this issue, it is important to acknowledge that complete elimination of reverse causality or other exogenous variables may not be achievable.

Time Horizon Limitations: The relatively short time span of our study further complicates our ability to precisely isolate the impact of new infrastructure over time. Factors such as non-linear effects and the duration of influence introduce complexities. To mitigate these challenges, we will make assumptions about a fixed timeframe for the effects to manifest and concentrate our analysis within these defined periods.

COVID-19 pandemic: The global COVID-19 pandemic introduces an additional layer of complexity to our analysis, potentially acting as a confounding factor. The pandemic has led to unprecedented changes in urban dynamics, including altered commuting patterns, shifts in housing preferences, and economic disruptions. These changes may confound the relationship between transportation infrastructure development and displacement, as the pandemic-induced effects could overshadow or interact with the anticipated impacts of new bike lanes and other projects.

9. Recommendations for Future Analysis

As we discussed in the limitations section, the scope of this project is limited to the scope of our available resources; we used only data on selected neighborhoods in Boston, in the years with available data. Our findings can and should be complemented by additional data, in other neighborhoods, cities, and years. Additionally, our analysis has been entirely quantitative. A supplementary qualitative research could enhance and improve the analysis and the accuracy of our results. For example, we had to use a few assumptions regarding the *relevance* of bike lanes on residents (e.g. for creating our "buffered" tract areas"); a survey asking people about their transportational habits and preferences can lead to better accuracy in this regard.

In case we would see some indications of a causality effect, a future step would be running a similar diff-in-diff in the opposite direction - for estimating whether changes in socio-economic variables are associated with future infrastructure changes. This is not necessary for this project, since our "step 1" was sufficient to not reject the null hypothesis; however, future analysis could benefit from this sort of analysis to help identify potential reverse causality, if it exists.

10. References

Ferenchak, N. N., & Marshall, W. E. (2021). Bicycling facility inequalities and the causality dilemma with socio economic/socio demographic change. Transportation research part D: transport and environment, 97, 102920. https://www.sciencedirect.com/science/article/pii/S1361920921002194?via%3Dihub

Geoghegan, Peter. (2016). “Blame it on the bike: does cycling contribute to a city's gentrification?” *The Guardian.* https://www.theguardian.com/cities/2016/oct/05/blame-bike-cycling-contribute-city-gentrification

LivableStreets Alliance (2023). “DOT Greenway”. *EmeraldNetwork.info*. https://www.emeraldnetwork.info/dot_greenway

McKim, Jenifer and Lugli, Stefania. (2023). 'They Want To Push Us Out': Mattapan Renters Fear Eviction As New Rail Stops Drive Rent Increases. *WGBH*. https://www.wgbh.org/news/local/2020-12-14/they-want-to-push-us-out-mattapan-renters-fear-eviction-as-new-rail-stops-drive-rent-increases

U.S. Census. (2023). Glossary. *United States Census Bureau*. https://www.census.gov/programs-surveys/geography/about/glossary.htm